DOCUMENT RESUME

ED 446 105 TM 031 848

AUTHOR	Ediger, Marlow
TITLE	Keys to Interpreting Research Results for Teachers and School Administrators.
PUB DATE	2000-00-00
NOTE	12p.
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Educational Research; Elementary Secondary Education; *Evaluation Methods; *Research Utilization; Scores; Test Results

ABSTRACT

This guide contains suggestions for interpreting educational research results and, indirectly, makes suggestions for research report writers. The purpose of the study and hypotheses should be stated clearly. Several approaches to reporting statistical results are outlined, and some suggestions are made for evaluating the results of cause-and-effect studies, correlational studies, and the results of questionnaire studies. It is important that those who use educational research be able to recognize potential weaknesses, and this is particularly important for research studies related to student learning in schools. Also discussed is the current trend toward considering context in the assessment of student achievement. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 446 105

Keys to Interpreting Research Results for Teachers and School Administrators

Marlow Ediger

TM031848

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as
received from the person or organization
originating it.
Minor changes have been made to
improve reproduction quality.
Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

Marlow Ediger

KEYS TO INTERPRETING RESEARCH RESULTS FOR TEACHERS AND SCHOOL ADMINISTRATORS

Much research is quoted in educational journals and speeches at teacher education conventions to support a point of view in teaching and learning. Little is done seemingly to determine the quality of the research study. There are many variables that need to be accounted for when doing research. When completing my descriptive survey master's thesis at Kansas State Teacher's college (1960), one delimitation stated that the study was delimited to one county in Kansas and not to other counties or states in the nation. When completing my doctoral dissertation at the University of Denver (1963), my experimental study stated for each hypothesis tested that "according to the Iowa Test of Basic Skills..." which indicated that the measurement instrument used stated a contextual conclusion. Thus, a different measurement device might have reached an alternative conclusion. Which ingredients should consumers of educational research have knowledge about?

Purpose of the Study

The purpose of the research study should be clearly stated. Vague, hazy statements need to be eliminated. Opinions arrived at, prior to conducting the study, need to be eliminated. The reason research is conducted is to attempt to become increasingly objective. "Increasingly" means that perfect objectivity is impossible since human beings are conducting the study. Also, subjects used in the study are humans and not automatons. When dealing with people then, the research deals with persons of different socio-economic levels, intelligence(s) possessed, achievement capabilities, motivation, and reasons for learning, among others. Consumers of research need to remember that human beings are different one from the other and "one size does not fit all." If the purpose of a study is to determine whether students do better with phonics versus whole language approaches in beginning reading instruction, objectively, the researchers should attempt to determine which is better, realizing that many variables among learners in the study are in evidence.

Testing an Hypothesis

Hypotheses need to be written with clarity. Supposing a research states the following objective: Students in the experimental group with whole language reading instruction will achieve higher than the control group with phonics instruction, at the .05 level. The level of significance could be increased in stringency from the.05 to the .01 level. The significance level may even be made more stringent to the .001 level.

Thus, the consumer or reader of research needs to be aware of the statistical procedure stated for the level of significance tested in the hypothesis. What are selected issues and problems when levels of significance between groups are to be measured?

1. the results may be very close to the .05 level so that the consumer of research needs to notice if the differences in achievement between the two groups, however, is important.

2. there will always be students who achieve less in the experimental or the control group, which ever is close to or significant at the .05, .01. or .001 level, than they would in the opposite group. This can be noticed by observing test scores of individuals from pretest to post-test results.

3. the formulas used to determine levels of significance are too difficult for many to understand. Thus, these formulas may well need to be accepted upon "faith."

4. there is considerable disagreement upon which standardized test should be used to measure pretest/ post-test or post-test only, achievement of the experimental group versus the control group. Each standardized test contains different test items as compared to the others.

5. there is also disagreement if the T-test or the F-test should be used as a statistical procedure when measuring significant differences between two groups. When three groups are being compared, the F-test is used only.

6. experimental studies are based upon the mean or average achievement in terms of gains between the experimental (the new approach in teaching) versus the control group (using the traditional procedure of instruction). A mean or average may be like "sitting on a cake of ice and having one's feet in boiling water and saying the average between the two is fine." Within the so called average, students differ much in achievement from high to low. Thus, for example, a low achieving student may do well on a pretest- post test basis in the research study whereas a talented learner may not do as well. The measurement instrument used then determines who is achieving well or a lack thereof. A different standardized test might measure quite differently.

7. studies comparing achievement of the experimental versus the control group are called cause and effect studies. Thus, the new approach in teaching is the cause for whatever happens which is the effect, from pretest to post test achievement of a group. And yet, there are so many variables that may affect how well students do in the experimental group where the new approach is used in instruction. Various causes for the effect include higher interest in one group as compared to the other set of students, as well as the halo effect for a new approach used in teaching. All eyes in the school and community may then be upon the new approach, whereas the traditional procedure of

instruction receives short shrift from observers. Another very important variable is that teachers for the experimental versus the control group can vary much in teaching skills and abilities.

8. random sampling procedures are recommended to be used in experimental studies. Thus, for either group -- the experimental versus the control, it may not be possible to use random sampling procedures to establish initial equality between the two groups when the study begins. After all, in most schools, students are in intact groups. These groups were determined, not for purposes of doing research in most cases, but to implement a certain philosophy of instruction. Intact groups then are not based on random sampling, such as using a table of random numbers, to determine which students go into the experimental group with the new approach in teaching and which go into the control group with the traditional approach of instruction. A quasi-experimental design may then be used, but it does not have the prestige that random sampling has to determine membership of students in the experimental and control groups. With quasi-experimental deigns, analysis of covariance is frequently used so that both groups start out at the same point in initial achievement. This is necessary so that one group does not have a head start over the other in the research study.

9. the theory of random sampling has its many weaknesses in that a standard error of the mean (SE Mean) formula is available to indicate how one sampling as compared to another sampling from the same set of students may be quite different. The larger the SE Mean, the greater the differences are in achievement when comparing one sample with another from a given set of students (See Kitchens, 1987).

10. standardized tests used for pertest-post test measurements have a built in procedure to spread students out in achievement from the 99th percentile to the first percentile. Thus, a spread of scores is desired by writers of standardized tests in order to obtain means, standard deviations, and a normal distribution for a bell shaped curve.
Many states in the United States emphasize the use of criterion referenced tests (CRTs) in which there are predetermined objectives for students to achieve. These objectives are then available to teachers to align the latter's learning opportunities with the chosen ends of instruction. Thus, the spread of scores of CRTs from high to low may not be nearly as great as compared to results from standardized testing. CRTs are based on a different philosophy of measurement as compared to standardized, also called norm referenced testing (Ediger, 2000, 503-505).

Correlational Studies

Correlational students also provide numerical results as do cause and effect studies. Correlational studies attempt to show a relationship

between/among variables. For example, if there are two variables, Intelligence Quotient (IQ) and academic achievement scores, what is the relationship between the two? One does not cause the other, but may be related, in degrees. Thus, a test is given to a select number of students for each of the two measures -- intelligence and academic achievement. The results are then compared with a correlational study. When looking at rank order from high to low on the individual results from IQ testing and doing the same thing from the achievement test, were individual students high on both measures, or low on both measures? The higher the correlation, the more closely related will be the result that students individually were at the same ranking for both measures -- IQ and achievement test results. Thus, for example, in using a small number of students, the following rank order, from high to low, shows a perfect 1.00 positive correlation between IQ test results and achievement test results:

IQ test results	Achievement test results
1. Bill	1. Bill
2. Sue	2. Sue
3. Martin	3. Martin
4. Adel	4. Adel
5. Adelbert	5. Adelbert
6. Jamielle	6. Jamielle
7. Addis	7. Addis
8. Babbettee	8. Babbettee
9. Sam	9. Sam
10. Otto	10. Otto

What happens in a correlation if the above named first column of names stays as is and the second column is inverted? There would then be negative correlation of - 1.00. This would show that there is a negative correlation between IQ and academic achievement. Low or no correlation would indicate the rank order of the above in the second column being in a random order or some place between the positive and negative correlation. What can correlations show that may have value for educators?

1. the relationship of phonics achievement versus meaning obtained from reading using whole language approaches. Suitable tests with high validity and reliability need to be used here for both measures.

2. rank order correlations may be used when data on student ranking is available, but their actual test scores are not available. If the rank order is very close, separating several students by a few raw score points, then these differences might not be important. Thus, the differences would be insignificant. In commencement ceremonies for high school graduation, the top five students names may be given in

rank order. When the author's achievements were needed to be included to be accepted for the Sate Farmer Degree in the Future Farmers of America (FFA) organization, open to two per cent of its members, his vocational agriculture instructor had rank order data available only, on senior level high school class academic achievement. Rank order was asked for in the FFA application to ultimately be selected to receive the State Farmer degree.

Raw scores may also be used, not rank order only, to develop correlations.

3. appropriate ranking is dependent upon the accuracy of measurement data used such as standardized test scores with their validity and reliability, or grade point average (gpa), in school. It does not tell how well a person will do in life in society. No one is tested in society with a paper/pencil test indicating how proficient a person is at the work place. Direct observation, among other procedures, are used to assess worker proficiency. How well a person does on the job, rather than on a paper/pencil test, is of utmost importance.

4. correlational studies attempt to determine the relationship between/among variables. They do not state the cause(s) of these relationships. Thus, it would be beneficial to know what causes scores to go up or down. This is especially true in a measurement movement era. Apparently, much time is spent by teachers to up test scores due to pressure from society. Testing appears to be the name of the game!

5. more studies need to be made pertaining to the correlation of test scores, grade point averages, among other variables, and success later at the work place. Multi-variate analysis is involved here with multiple correlations. An additional problem here is how to ascertain and measure success on the job.

Questionnaires and Educational Research

Frequently, questionnaires are developed and mailed to a random sampling of respondents.The results can fail if the items on the questionnaire are vaguely written, making for misunderstandings by the respondent. A second problem generally is a low level of return from respondents. Sometimes, a 20 % return is reported in a research study. That is a very low level of return from respondents. The questionnaire was mailed to determine the feelings of the total number of respondents, not the 20% only. It is impossible to secure a 100% return to the mailed questionnaire. Researchers, generally, find an 80% return to be acceptable.

Third the developer/writer of the questionnaire can select a sample of possible respondents who definitely are not representative of the total population. Thus, a biased sampling of respondents provides very distorted findings in a questionnaire. A newsreporters asking the

first twenty teachers he/she meets how they feel about merit rating has met up with a biased sample of respondents. A random sampling was then not chosen for the study. A Table of Random numbers may be used whereby everyone in the defined set has an equal chance to be chosen in becoming a member of the research study.

Fourth, a margin of error is always present in questionnaire results. The margin of error represents a range from which returns may be ascertained. Thus, a margin of error of two points may mean that if 52%, for example, are going to vote for candidate A in the general election, the range could really be from 50% to 54%. There are weaknesses in questionnaire results and the 52% actual results may vary two points from 52% minus two =50%, to 52% plus two per cent = 54%. The race to win may then be in a dead heat with 50% of the votes possible for a candidate with the margin of error included.

Fifth, the accuracy of a questionnaire is dependent upon the honesty of respondents. In rating the university instructor's quality of instruction at the end of a course, students can be in a hurry to leave the classroom. Little attempt then is put forth in carefully reading and conscientiously responding to each item on the questionnaire to evaluate the quality of instruction. A university student may also feel this is the time to "get even" with the instructor by marking the quality of instruction in a negative manner. Weaknesses in questionnaire development need to identified and remedied.

When pointing out weaknesses in doing research, does this mean it should be abolished? No, definitely not Research needs to be done continually with the reader knowing weaknesses inherent therein. Then too, researchers need to refine their methods and procedures in conducting research. Educators who summarize research studies need to include only those which meet criteria of excellence. Shoddy research should not be footnoted in a research article for an educational journal. The honesty of the researcher needs to be unquestionable!

Research Results on Student Learning in Schools

There is considerable criticism of student achievement in public schools. Much of the criticism is opinion from the business world and the lay public. Quality research is hard to come by and is expensive to conduct. Thus the controversy of phonics versus whole language in reading instruction has not been resolved. Why, there are many weaknesses in research that has been conducted in this area. Which are selected areas of weaknesses that need identification?

1. much of the research is short term in nature, perhaps one school year or less in duration. Rather longitudinal research needs to be in evidence. Thus, the research needs to be carried on over a period of years. Why is this important? Students receiving the new approach in

teaching may do well on the primary grade levels, but the newness of the approach so often wears off as students progress through the intermediate and beyond grade levels.

2. primary grade students are compared to intermediate grade students in selected research studies. This is a biased comparison made pertaining to which group does better in achievement using a new approach in teaching and a control group with traditional procedures. Why? Primary grade students undergo many more changes as compared to intermediate grade students. From a helpless infant at birth to the end of the first grade, many students are reading early or later primary grade reading materials. Intermediate grade students do not change that rapidly; however, the changes are definitely there with many becoming quite proficient readers by the end of the intermediate grade level (See Piaget, 1950).

3. selected groups are left out of the research. For example, to have a new approach in teaching look good, a researcher leaves out mentally retarded students from the experimental group. The control group has students therein from all ability levels and are left intact. It is no wonder that the experimental group will show greater gains from the pretest to the post test. When international comparisons are made among nations such as the Third international Mathematics and Science Study (TIMSS), who was included in each nation's students to show which has the "best" educational system? If a nation has more of the cream of the crop students in their comparison with other nations, they are bound to show "superior" results.

4. there are reports of teachers and administrators changing student test scores to achieve at a higher rate, as demanded by state and city departments of education. Punishments for low school test scores include educational bankruptcy laws, withholding funding, and publishing report cards of low achieving schools, among others.

5. standards may be lowered within a state so that a higher percentage of high school graduates result. With lowering the bar and nothing being published pertaining to why more are graduating, the involved state then may look better in the eyes of the public with their same "higher" standards.

6. standards may also be raised considerably so that few students pass the exit test. The state of Virginia set standards so high that 98% of the schools failed their state wide test, while 91% of these same students failed the test the second time (Bracey, 2000). Why are standards set excessively high? The following are possible: to make public schools look bad, to pressure teachers to have students achieve at a higher rate, as well as a lack of knowledge and skills in test development.

7. no control group is used in the study to make comparison with the experimental group. If there is no control group with the traditional

approaches in teaching, then comparisons cannot be made. If a control group had been there, they might have done better than the experimental group using the newer procedure of instruction.

8. no experimental group is pure in using their methods of teaching. For example with whole language approaches in reading instruction, how many teachers use no phonics? Or, in phonics methodology, how many use a pure systematic phonics approach in teaching? The chances are that neither uses a "pure" approach. Even if these two terms are clearly defined -- whole language versus phonics, the chances are that either approach is used in degrees.

9. teacher skills and abilities vary much among teachers, from those students in the experimental group as compared to those in the control group.

10. external validity is lacking. The conditions of the experimental group with the new method of instruction may have a completely different setting as compared to the classroom of the reader of the research.Thus, the reader of the research may have the following classroom conditions: thirty five students as compared to the experimental group in the research having twenty in the classroom; the only mentors are classroom teachers as compared to available paid mentors in the experimental group; balance in the curriculum with all curriculum areas being adequately emphasized, as compared to a heavy emphasis placed upon reading with reading test scores going up much for the latter but not for the former; few instructional materials available as compared to adequacy in terms of diverse teaching supplies to provide for individual differences; inadequate school funding as compared to being in a high socio-economic area for teaching students (Ediger, 2000, Chapter Eight).

Contextualism in the Curriculum

A rather recent innovation challenging the testing and measurement movement is contextualism to indicate student achievement and progress. Contextualism emphasizes assessing student achievement within the ongoing lesson and unit of study being pursued. In some ways, contextualism as a philosophy of assessment has always been used by teachers in evaluating student achievement (See Ediger, 1997-1998, 56-60). Contextualism does not

1. advocate using state mandated objectives and tests in teaching learners.

2. stress individuals external to the local classroom writing tests for local students to take to ascertain achievement in the different academic areas.

3. proclaim numerical scores to show student achievement. Numerical results emphasize percentiles, standard deviations, quartile deviations, and stanines, among other standard scores.

4. separate assessment from teaching and learning activities.
5. compare schools, school systems, and individual students in achievement with the use of report cards.

Contextualism emphasizes assessing student achievement as being ongoing involving learner products and products (Ediger, 1995, 1-11). Thus within a learning opportunity, the teacher as needed may diagnose and assist the learner in

1. writing prose and poetry.
2. reading for a variety of purposes.
3. discussing what has been read.
4. helping a committee or individual to develop an oral report.
5. guiding students to become better listeners.
6. working with students on a science or social studies project.
7. explaining a task at a learning center for cooperative work or individual endeavors.
8. demonstrating how to take part in a creative dramatics or formal dramatization activity.
9. use of student/teacher planning to determine objectives and learning opportunities in an ongoing unit of study.
10. leading a group of learners in a problem solving experience.

All of the above named responsibilities emphasize curriculum development and assessment within the classroom. External people such as writers of tests and state department of education personnel were not mentioned. Does this mean external persons have no role in curriculum improvement? No. There is room for flexible state mandated objectives and tests. How should external personnel be involved?

1. determine relevant, carefully chosen objectives of instruction, not minutia.
2. take time to develop tests that have been pilot studied and revisions made before being used in the public schools.
3, use test results to improve the curriculum, not to ridicule the public schools with report cards and other devious devices.
4. constructively assess the public schools with quality remedies for improvement, but not use put downs.
5. be positive in presenting information to the lay public on the quality of public school education.
6. assist schools to place qualified teachers in each classroom, not emergency certified substitutes.
7. work together with the public schools in curriculum improvement and not be at loggerheads.
8. eliminate negative reports about the public schools unless there are constructive, feasible suggestions, as well as financial aid if needed, in working toward improvements.

9. make efforts for all students to achieve well and to support good public schools in order to provide for an educated citizenry as well as for productive workers at the work place, as ultimate goals.

10. help participants in society to work together well for the good of the learner.

References

Bracey, Gerald W., "Research," Phi Delta Kappan, 82 (3), 92.

Ediger, Marlow (1963), The Influence of the Student Teacher on the Pupil, Academically and Socially, In Selected Elementary Grades. A doctoral dissertation presented to the Graduate College, University of Denver.

Ediger, Marlow (1960), The Status of the Public School Teacher of Marion county, Kansas, 1959-1960. A master's thesis presented to the Department of Education and the Graduate Council, Kansas State Teacher's College of Emporia, Kansas.

Ediger, Marlow (1997-1998), "Portfolios and the Social Studies," Journal of the Illinois Council for the Social Studies, 56-60.

Ediger, Marlow (1995), "The Psychology of Learning and the Teacher," Philippine Education Quarterly, 23 (2), 1-11.

Ediger, Marlow (2000), "Choosing Evaluation Procedures," Education, 120 (3), 503-505.

Ediger, Marlow (2000), Teaching Mathematics Successfully. New Delhi, India: Discovery Publishing House, Chapter Eight.

Kitchens, Larry J., Exploring Statistics. St. Paul, Mn.: The West Publishing Company.

Piaget, Jean (1950), The Psychology of Intelligence. New York: Harcourt Brace Jovanovich, Inc.